aukcar.ac.uk
@AUKCAR

PhD studentship at the
Asthma UK Centre for Applied Research (AUKCAR):

Are anonymised databases truly anonymous? An Introduction

Ref: AUKCAR-17-01a

This work is funded by CMVM/UoE. This work is carried out with the support of the Asthma UK Centre for Applied Research [AUK-AC-2012-01].

By Aryelly Rodriguez
22 NOV 2017

Asthma UK Centre for Applied Research

THE UNIVERSITY of EDINBURGH

Background

Demands from funders, regulators and/or publishers to share clinical trials data to help with:

- Out of study scope questions
- Further exploratory analysis of outcomes
- Validation and transparency
- Analysis of Individual Participant Data (IPD) in meta-analysis
- Development of new methodologies
- Reducing costs and increase efficiency of health care
- Minimise the bias of positive reporting

Importance

- The importance and significant benefits of clinical data sharing have been well documented by many researchers in such as Gøtzsche(1), Packer(2), Al-Shahi et al.(3), Pisani et al. (4) and Bertagnolli et al. (5) just to cite a few:

Gøtzsche *Trials* 2011, **12**:249
http://www.trialsjournal.com/content/12/1/249

TRIALS

COMMENTARY Open Access

Why we need easy access to all data from all clinical trials and how to accomplish it

Peter C Gøtzsche

Saturday 28 October 2000

BMJ

Using patient-identifiable data for observational research and audit

Overprotection could damage the public interest

Across the world rapid changes in the law, technology, and society are reshaping the way identifiable information about patients is handled. In Britain, doctors' longstanding common law duty of confidentiality to their patients has been

because patient-i
duplication and t
recent legal ruling
without consent f
constitute a breac

Beyond open data: realising the health benefits of sharing data

Accessible data are not enough. We need to invest in systems that make the information useful, say **Elizabeth Pisani and colleagues**

Data sharing: lessons from Copernicus and Kepler

They did it, and so must we—quickly, reliably, and respectfully

Milton Packer *distinguished scholar in cardiovascular science*

Asthma UK Centre for Applied Research

asthma UK

THE UNIVERSITY of EDINBURGH

Importance (cont.)

"interests of the patients must override commercial interests"(1)

"A moral imperative"(1)

"Respect for trial participants who often run a personal and unknown risk by participating in trials requires that they - and therefore also the society at large that they represent - be seen as the ultimate owners of trial data"(1)

"National and supranational legislation is needed to make data sharing happen as guidelines and other voluntary agreements do not work"(1)

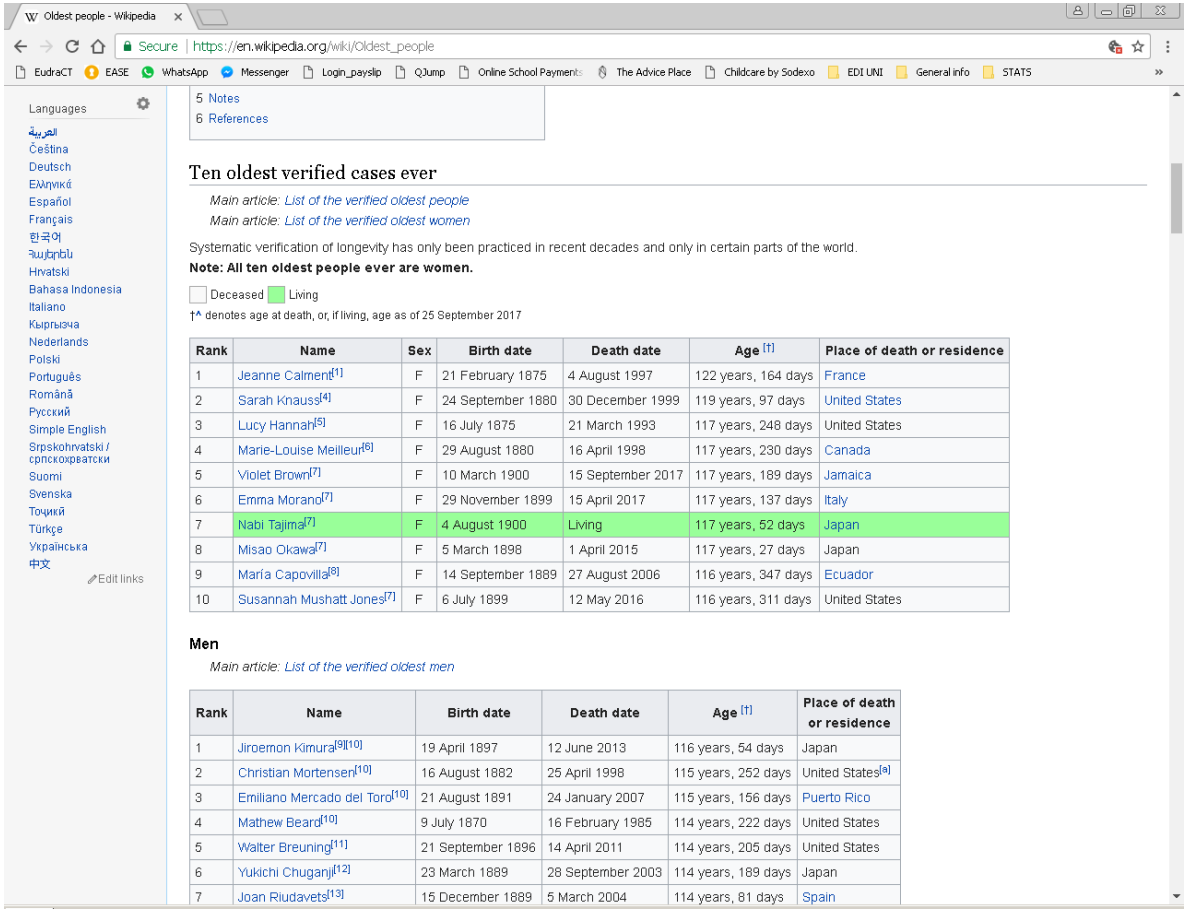
"Now look at the medical literature in the 21st century. We no longer publish our data; instead, we present truncated summaries in the hope that readers will believe our conclusions without seeing the raw observations…It is no wonder that many clinicians have stopped reading the medical literature"(2)

Current situation

- "Concerns that patient confidentiality and consent may be breached are often cited by researchers as a reason for not sharing data"(4)
- "Threat of data parasites" (4)
- Presence of overly restrictive internal governance (timelines and quality of data)
- Fear of the public's opinion

Current situation (cont.)

https://en.wikipedia.org/wiki/Oldest_people



Oldest people - Wikipedia

Secure | https://en.wikipedia.org/wiki/Oldest_people

EudraCT EASE WhatsApp Messenger Login_payslip QJump Online School Payments The Advice Place Childcare by Sodexo EDI UNI General info STATS

Languages

العربية Čeština Deutsch Ελληνικά Español Français 한국어 Հայերեն Hrvatski Bahasa Indonesia Italiano Кыргызча Nederlands Polski Português Română Русский Simple English Srpskohrvatski / српскохрватски Suomi Svenska Тоҷикӣ Türkçe Українська 中文

Edit links

5 Notes
6 References

Ten oldest verified cases ever

Main article: List of the verified oldest people
Main article: List of the verified oldest women

Systematic verification of longevity has only been practiced in recent decades and only in certain parts of the world.
Note: All ten oldest people ever are women.

Deceased Living

†^ denotes age at death, or, if living, age as of 25 September 2017

Rank	Name	Sex	Birth date	Death date	Age [†]	Place of death or residence
1	Jeanne Calment[1]	F	21 February 1875	4 August 1997	122 years, 164 days	France
2	Sarah Knauss[4]	F	24 September 1880	30 December 1999	119 years, 97 days	United States
3	Lucy Hannah[5]	F	16 July 1875	21 March 1993	117 years, 248 days	United States
4	Marie-Louise Meilleur[6]	F	29 August 1880	16 April 1998	117 years, 230 days	Canada
5	Violet Brown[7]	F	10 March 1900	15 September 2017	117 years, 189 days	Jamaica
6	Emma Morano[7]	F	29 November 1899	15 April 2017	117 years, 137 days	Italy
7	Nabi Tajima[7]	F	4 August 1900	Living	117 years, 52 days	Japan
8	Misao Okawa[7]	F	5 March 1898	1 April 2015	117 years, 27 days	Japan
9	María Capovilla[8]	F	14 September 1889	27 August 2006	116 years, 347 days	Ecuador
10	Susannah Mushatt Jones[7]	F	6 July 1899	12 May 2016	116 years, 311 days	United States

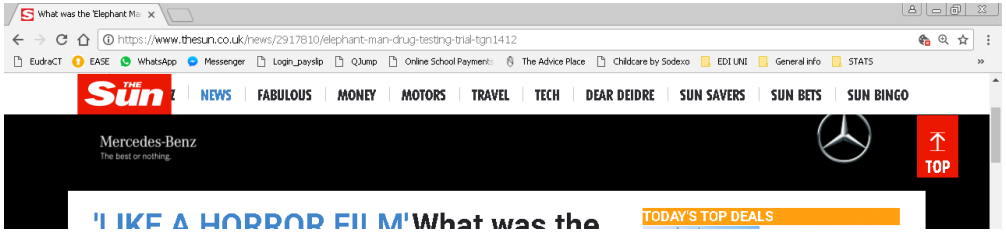
Men

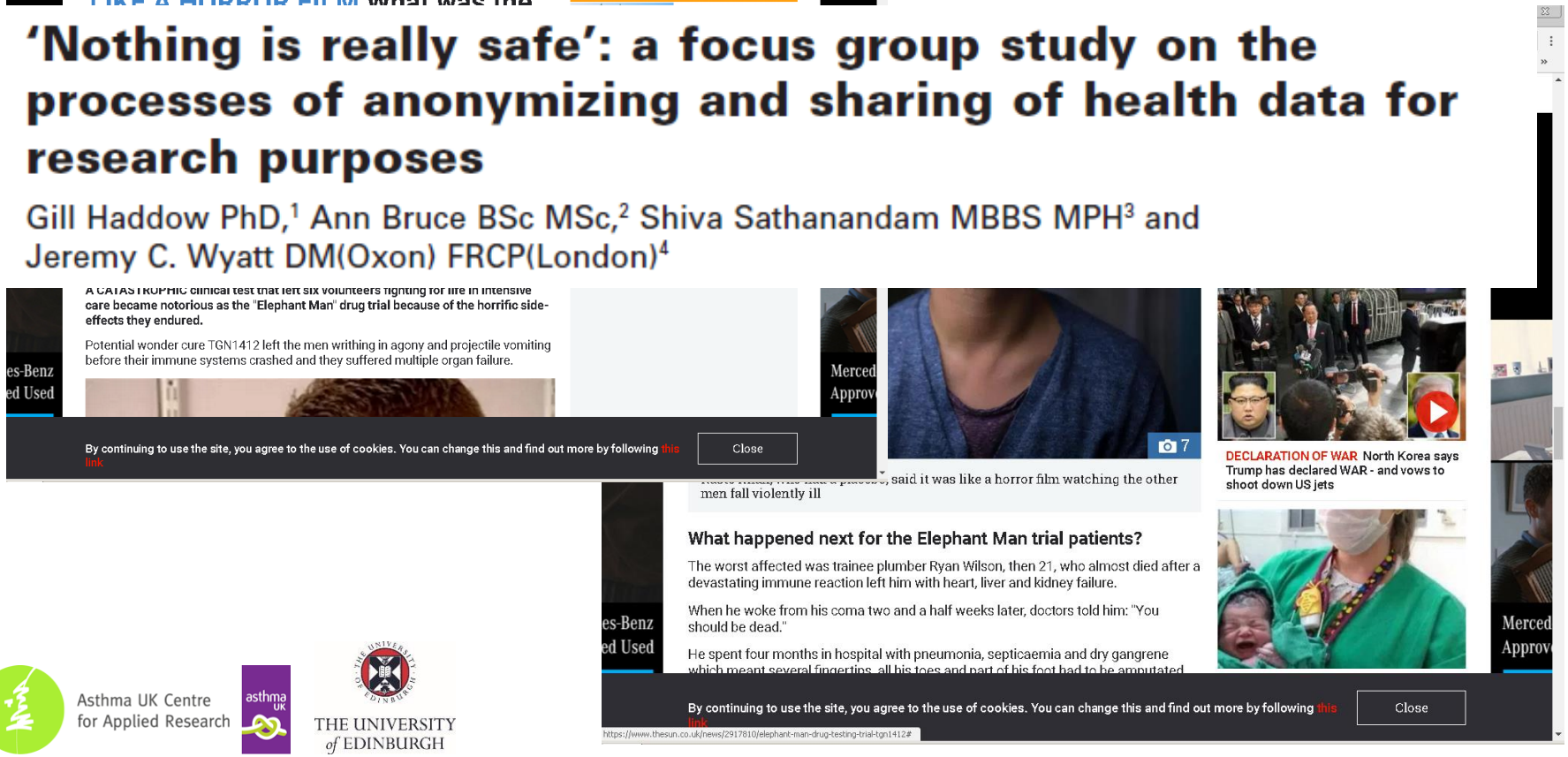
Main article: List of the verified oldest men

Rank	Name	Birth date	Death date	Age [†]	Place of death or residence
1	Jiroemon Kimura[9][10]	19 April 1897	12 June 2013	116 years, 54 days	Japan
2	Christian Mortensen[10]	16 August 1882	25 April 1998	115 years, 252 days	United States[a]
3	Emiliano Mercado del Toro[10]	21 August 1891	24 January 2007	115 years, 156 days	Puerto Rico
4	Mathew Beard[10]	9 July 1870	16 February 1985	114 years, 222 days	United States
5	Walter Breuning[11]	21 September 1896	14 April 2011	114 years, 205 days	United States
6	Yukichi Chuganji[12]	23 March 1889	28 September 2003	114 years, 189 days	Japan
7	Joan Riudavets[13]	15 December 1889	5 March 2004	114 years, 81 days	Spain

Current situation (cont.)

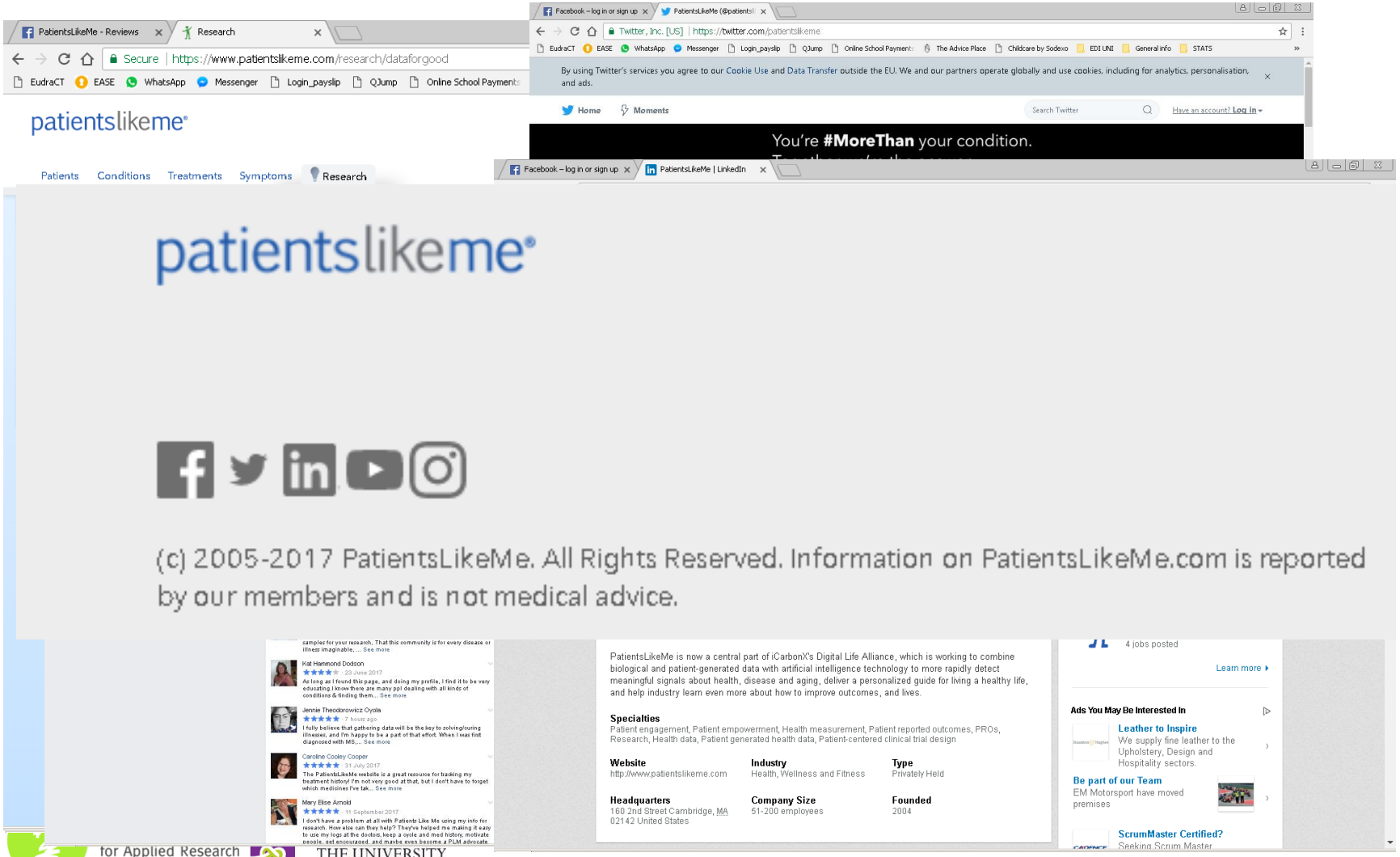
https://www.thesun.co.uk/news/2917810/elephant-man-drug-testing-trial-tgn1412





Current situation (cont.)

https://www.patientslikeme.com/



“The Issue” for this project

Are patients at risk of being identified under the current methods / guidelines used for anonymisation at CTUs?

The PhD project aims and objectives are:

a. Investigate whether individual participants can be identified from a range of datasets that have been anonymised and made available for sharing

b. Identify factors that could increase the risk of re-identification of a dataset

c. Develop evidence-based recommendations on anonymization techniques and data security

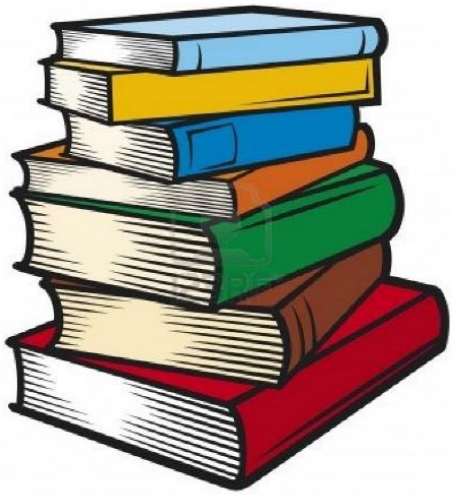
d. Further explore researchers, patients and public perception of the sharing of clinical research data under the scenarios identified by this research

Literature Review

A systematic literature review is currently being executed with the goal to investigate:

- The current trends and practices among researchers/ organisations regarding anonymization techniques and policies for data sharing
- The potential re-identification methods that can be used on anonymised IPD
- The strategies that are being used to protect anonymised and published data against re-identification

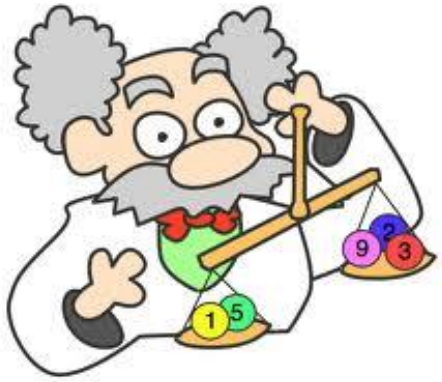
The literature review is including academic literature, reviews/reports in national media (UK) and social media such as blogs, Twitter and Facebook



Quantitative phase

Data collection: A selection of datasets that have been anonymised and made available for sharing will be obtained from the Edinburgh Clinical Trial Unit, Asthma UK, other Clinical Trial Units registered with UK-CRC, private sponsors such as GSK and from peer reviewed medical journals

Analysis: All collected datasets will be crosschecked against any relevant re-identifying source using publicly available information (including study publications and web resources) to investigate whether any small groups or individuals can be identified.
Also, the datasets will be assessed to determine how usable they are for further research.
Finally, characteristics of the datasets (e.g. size of the study, public vs privately funded, rare vs common conditions) will be recorded to assess the potential risk of re-identification



Qualitative phase

Canvas academic community, patients and public opinion about:

- The requirement for sharing IPD, including recognition of possible motives for an individual or organisation to identify individual participants
- Strategies to protect anonymity of patients if their study becomes newsworthy

This would be delivered by using questionnaires, focus groups and interviews

Overall Schedule for key milestones

Year	Month	Literature Review	Quantitative phase	Qualitative phase
1-2	0-10	Perform review	Issue requests and collect datasets	Identification of stakeholders and tool development for data collection
	12	Publish results		
	14-22			Send out questionnaires
	24			Collect questionnaires, hold focus groups and interviews
3-4	26-44			
	36		Summarise progress	
	38-46		Test datasets	Code and Analyse responses
	48		Generate and Publish results	
5-6	50-58		Characterise datasets and determine risk factors	
	60		Generate results	Generate and Publish results
	62-72	Write PhD dissertation		

Many thanks to my funders

Asthma UK Centre
for Applied Research

THE UNIVERSITY
of EDINBURGH

And supervisors:

Prof S Lewis (University of Edinburgh) steff.lewis@ed.ac.uk
Prof C Weir (University of Edinburgh) Christopher.Weir@ed.ac.uk
Prof S Eldridge (Queen Mary University of London) s.eldridge@qmul.ac.uk

References

(1) Gøtzsche PC. Why we need easy access to all data from all clinical trials and how to accomplish it. Trials. 2011;12:249. pmid:22112900 Why we need easy access to all data from all clinical trials and how to accomplish it

(2) Packer Milton. Data sharing: lessons from Copernicus and Kepler BMJ 2016; 354 :i4911

(3)Al-Shahi, R. & Warlow, C. (2000) Using patient-identifiable data for observational research and audit. BMJ, 321 (7268), 1031–1032.

(4) Pisani Elizabeth, Aaby Peter, Breugelmans J Gabrielle, Carr David, Groves Trish, Helinski Michelle et al. Beyond open data: realising the health benefits of sharing data

BMJ 2016; 355 :i5295

(5)Monica M. Bertagnolli, M.D., Oliver Sartor, M.D., Bruce A. Chabner, M.D., Mace L. Rothenberg, M.D., Sean Khozin, M.D., M.P.H., Charles Hugh-Jones, M.D., David M. Reese, M.D., and Martin J. Murphy, D.Med.Sc., Ph.D. Advantages of a Truly Open-Access Data-Sharing Model. N Engl J Med 2017; 376:1178-1181, March 23, 2017

DOI: 10.1056/NEJMsb1702054

(6) Haddow, G., Bruce, A., Sathanandam, S. and Wyatt, J. C. (2011), 'Nothing is really safe': a focus group study on the processes of anonymizing and sharing of health data for research purposes. Journal of Evaluation in Clinical Practice, 17: 1140–1146. doi:10.1111/j.1365-2753.2010.01488.x

aukcar.ac.uk

@AUKCAR

Thank you!

Any questions?

Asthma UK Centre for Applied Research